DOCUMENT RESUME

ED 388 691 TM 023 934

AUTHOR Sheehan, Janet K.
TITLE A Comparison of the Type I Error and Power of Selected MANOVA Simultaneous Test Procedures.
PUB DATE Apr 95
NOTE 44p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Analysis of Variance; Comparative Analysis; Monte Carlo Methods; *Multivariate Analysis; *Robustness (Statistics)
IDENTIFIERS Hotellings t; Pillais Trace; *Power (Statistics); Roys Largest Root; *Type I Errors

ABSTRACT

A Monte Carlo study was conducted using the Statistical Analysis System IML computer program to compare the multivariate analysis of variance (MANOVA) simultaneous test procedures of Roy's Greatest Root, the Pillai-Bartlett trace, the Hotelling-Lawley trace, and Wilks' lambda, in terms of power and Type I error under various conditions, including violations of MANOVA assumptions. The Type I error rates of moderately-restricted contrasts in simultaneous test procedures following a significant omnibus MANOVA were robust to violations of MANOVA assumptions, such that the actual alpha remained below the nominal alpha. However, the power of even Roy's Greatest Root is unacceptably low in moderately restricted contrasts under most conditions. Therefore, the results of this study do not generally support using moderately restricted contrasts to follow-up significant MANOVA tests, unless the number of dependent variables is limited to two, or the noncentrality structure is known to be concentrated in one group and one variable. (Contains 1 table, 12 figures, and 35 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

running head: MANOVA SIMULTANEOUS TEST PROCEDURES

A Comparison of the Type I Error and Power of Selected MANOVA Simultaneous Test Procedures

Janet K. Sheehan

Northern Illinois University

ED 388 691

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JANET K. SHEEHAN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This paper was presented at the Annual Meeting of the American Educational Research Association San Francisco, CA, April 18-22, 1995

TM023434

BEST COPY AVAILABLE

ABSTRACT

A Monte Carlo study was conducted using SAS-IML to compare the MANOVA simultaneous test procedures of Roy's Greatest Root, the Pillai-Bartlett trace, the Hotelling-Lawley trace, and Wilks' lambda, in terms of power and type I error under various conditions, including violations of MANOVA assumptions. The type I error rates of moderately-restricted contrasts in simultaneous test procedures following a significant omnibus MANOVA were robust to violations of MANOVA assumptions, such that the actual alpha remained below the nominal alpha. However, the power of even Roy's Greatest Root is unacceptably low in moderately-restricted contrasts under most conditions. Therefore, the results of this study do not generally support using moderately-restricted contrasts to follow-up significant MANOVA tests, unless the number of dependent variables is limited to two, or the noncentrality structure is known to be concentrated in one group and one variable.

A Comparison of the Type I Error and Power of Selected MANOVA Simultaneous Test Procedures

Multivariate analyses in educational and psychological research have become much more prevalent since the 1970s (Maxwell, 1992). Emmons, Stallings, & Layne (1990) surveyed sixteen years of research and determined that "The multivariate characteristic of the social science research environment with its many confounding or intervening variables has been addressed through the trend toward increased use of multivariate analysis of variance and covariance, multiple regression, and multiple correlation." (p.14).

Multivariate analysis of variance (MANOVA) is generally used to determine if there are group differences on a set of p variables. Post-hoc follow-up procedures are arguably more critical in the multivariate case than the univariate case. The omnibus MANOVA not only fails to delineate where the group differences occur, but also fails to describe on which variables these differences lie.

MANOVA Test Statistics

The four MANOVA test statistics: W, V, T, and R, combine the information from the s eigenvalues of the $\mathbf{HE}^{-1}$ matrix in different ways to test the multivariate hypothesis (Bray & Maxwell, 1985; Olson, 1976). Other test statistics based on these eigenvalues are inferior to at least one of these four statistics (Olson, 1976).

Wilks' lambda is the oldest multivariate test statistic, and is the most widely used (Tatsuoka, 1988). W is a function of the product of the s roots, or alternatively can be expressed as a ratio of determinants (Wilks, 1932).

$$W = \prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|}. \qquad (1)$$

Wilks' lambda is often recommended, because of its computational ease (Schatzoff, 1966). Moreover, W is conceptually easy to understand , because it is a ratio of determinants. Hence, W is a ratio of the generalized variance of the **E** matrix to the **T** matrix (**T** = total sum of squares and cross-products). Therefore, W decreases as the multivariate effect size increases.

Both the T and V multivariate test criteria are based on the trace of a matrix. T is the trace of the $\mathbf{HE}^{-1}$ matrix (Hotelling, 1931; Lawley, 1939).

$$T = \sum_{i=1}^{s} \lambda_i . \tag{2}$$

V is the trace of the $\mathbf{HT}^{-1}$ matrix or is equivalent to the following function of the $\mathbf{HE}^{-1}$ matrix (Bartlett, 1939; Pillai, 1955) .

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} . \tag{3}$$

Hence, V and T increase in size as the multivariate effect size increases. Further, it is known that W, V, and T are asymptotically equivalent in very large samples. Empirical results suggest that they may be considered equivalent when df_e is at least 10p times larger than df_h (Olson, 1976).

In contrast, R is simply a function of the largest root. R is the largest eigenvalue of the $\mathbf{HT}^{-1}$ matrix (Roy, 1945). R is a function of the $\mathbf{HE}^{-1}$ matrix as follows:

$$R = \frac{\lambda_1}{1+\lambda_1} . \tag{4}$$

R also increases as the multivariate effect size increases.

When df_h=1, all of these test statistics become a function of the first eigenvalue, and hence are all proportional:

$$T = \lambda_1, \quad R = V = \frac{\lambda_1}{1+\lambda_1}, \quad W = \frac{1}{1+\lambda_1}. \tag{5}$$

When df_h>1, the test criteria values diverge and conclusions based on them may differ. T, V, and W are more useful for detecting a noncentrality structure that is divided among the *s* roots; a diffuse structure. By comparison, R is the best choice for isolating a noncentrality structure that is located in one root; a concentrated structure. Empirical studies have supported this inferred relationship between the test statistics and the noncentrality structure. Schatzoff (1966) compared the relative sensitivities of six multivariate test criteria, including V, T, W, and R, under a variety of population structures. The population structures did not violate any of the multivariate assumptions. When the noncentrality structure was very diffuse, the sensitivity for detecting the population structure was ordered V>W>T>R. When the noncentrality structure was concentrated in one root, the sensitivity was reversed. R had the greatest ability to detect the population structure, and V had the worst ability to do so.

MANOVA Test Statistics and Violations of Multivariate Assumptions

Olson (1974) investigated the presence of kurtosis and variance-covariance heterogeneity on the power and robustness of six MANOVA test statistics, including R, T, W, and V. Olson confirmed the patterns Schatzoff found when the multivariate assumptions were upheld. However, Olson found that when the population structures had violations of these assumptions, the sensitivity patterns changed. Moreover, these four test statistics differed in robustness to violations of multivariate assumptions.

The ordering of empirical power remained R>T>W>V when the noncentrality structure was concentrated, whether or not multivariate assumptions were violated. However, the relationship of the power of the different test criteria observed by Schatzoff (1966) for diffuse noncentrality structures did not hold for some situations with violations of multivariate assumptions in one-way MANOVAs with equal n's. When multivariate normality was violated due to kurtosis, the difference in the power between V, T, and W was very small, and R was less powerful than all three (Olson, 1974). The power of W, V, T, and R usually decreased when the assumption of homogeneity of variance-covariance matrices was violated. The power for V was considerably lower than the other three statistics under some conditions when the assumption of homogeneity of variance-covariance matrices was violated (Olson).

In large samples with equal n's T and W have been shown to be robust to violations of variance-covariance homogeneity, however, samples with unequal n's were severely affected by variance-covariance heterogeneity, even in very large samples (Ito, 1969; Ito & Schull, 1964). In small samples the T, W, and R statistics were not robust to violations of the homogeneity of variance-covariance assumption, even with equal n's (Korin, 1972) (see Table 3). However, violations of the multivariate assumptions often had varying effects on the exceedance rates of the four test criteria. An important factor that affected exceedance rates was whether the contamination of the assumption violation occurred equally in all dimensions of the dependent variable set (low concentration of contamination), or whether the contamination occurred in one dimension of the dependent variable set (high concentration of contamination). A low concentration of contamination had more impact on exceedance rates than a high concentration of contamination. When positive kurtosis was present, all four of the test statistics were conservative; the ordering of exceedance rates among the test criteria was V>W>T>R.

Heterogeneity of variance-covariance had a liberal effect on the exceedance rates. In this case, R was the most liberal and the ordering of exceedance rates among the test criteria was R>T>W>V.

Olson (1974, 1976) recommended V for general use, because V was the least conservative in the presence of kurtosis, and the least liberal in the presence of variance-covariance heterogeneity. Although V tended to be least powerful when variance-covariance heterogeneity was present, Olson believed it had adequate power in most situations. Stevens (1979) disagreed with Olson's (1974, 1976) unilateral endorsement of the V statistic when violations of assumptions occur. Stevens recommended using T, W, or V for concentrated structures when variance-covariance heterogeneity is present.

All of these studies compared the power or robustness of multivariate test statistics for the omnibus test. Therefore, these recommendations are reasonable only if the prime concern is to detect an overall effect. However, there has been considerable interest in the multivariate literature in attempting to discern what variables and/or which groups contribute most to the multivariate significance.

Interpreting the Multivariate Effect

There are five general procedures that are used to further investigate a significant omnibus test in MANOVA: selecting subsets of variables through discriminant analysis, step-down analysis, two group comparisons, planned contrasts, and simultaneous confidence intervals (SCI's) or simultaneous test procedures (STP's). The first two of these are concerned with determining which criterion variables contribute most to the overall group differences. Either the structure coefficients or the discriminant function coefficients generated from the discriminant analysis can be used to aid in interpreting the combination of dependent variables that contribute to each discriminant function variate. However, as McKay and Campbell (1982) observe, selection methods based on

discriminant analysis are arbitrary. If discriminant function weights are used, they must be recalculated after every step of variable deletion to base further decisions on. Highly multicollinear variables can produce very unstable discriminant function coefficients and muddle the interpretability of the discriminant function variate. McKay and Campbell also point out that basing variable-deletion decisions on the values of the structure coefficients is not theoretically sound. Consequently, selecting variables by these methods often renders misleading information and may result in loss of ability to separate groups.

Another technique for determining the variables that contribute most to multivariate significance is step-down analysis (Bock, 1963; Roy, 1958). The dependent variables are first ordered according to theoretical importance. The highest priority variable is tested with a univariate ANOVA. The analysis then proceeds as an analysis of covariance. In each step the next highest-priority variable is tested with the higher-priority variables as covariates. When an insignificant F-statistic is generated, the analysis stops. The final subset of variables are all of the higher-priority variables that reached significance. This method is not feasible if the variables in the dependent set cannot be ordered *a priori*. A further consideration is that this method does not directly capture the root which may be of primary theoretical interest.

Another multivariate post-hoc technique compares pairs of groups on the set of variables using Hotelling's T^2 (Stevens, 1986). The significant multivariate test can subsequently be followed with univariate t-tests to determine which variables significantly contribute to the group separation (Stevens). This method has the advantage over previous methods that it examines both the independent and dependent variable set to tease out the significant multivariate effects. However, this method yet fails to fully address the multivariate question, because

it ultimately reduces to univariate tests and ignores the correlations among the dependent variables.

Planned multivariate contrasts are truly multivariate procedures that examine contrasts of the groups across composites of the dependent variables. A multivariate contrast, φ, is equal to $\mathbf{c}'\mu\mathbf{a}$. $\mathbf{c}$ is a k-element vector of contrast coefficients for the k groups; μ is the k p matrix of population means; and $\mathbf{a}$ is equal to a p-element vector of variate coefficients (Bird, 1975; Bird & Hadzi-Pavlovic, 1983). If the group contrast coefficients and variate coefficients can be specified before the analysis is conducted, then this is an *a priori* multivariate test procedure. Planned comparisons of this type can be tested as single degree of freedom F-tests (Harris, 1985, p. 103-105). Planned multivariate comparisons are preferred over multivariate post-hoc comparisons because of their greater power. However, their usefulness is limited to situations in which the researcher has a theoretical basis for a particular comparison on both the independent and the dependent variable set.

Multivariate Simultaneous Test Procedures

When it is desired to follow-up an omnibus MANOVA with post-hoc comparisons of a truly multivariate nature, simultaneous confidence intervals (SCI's) (Roy & Bose, 1953) or simultaneous test procedures (STP's) (Bird & Hadzi-Pavlovic, 1983; Gabriel, 1968; McKay & Campbell, 1982) can be used. Roy and Bose first described a multivariate SCI using Roy's Greatest Root. The multivariate contrast, $\mathbf{c}'\mu\mathbf{a}$, is estimated at the $1 - \alpha$ confidence level by the interval:

$$\mathbf{c}'\overline{\mathbf{X}}\mathbf{a} - \sqrt{\frac{\mathbf{c}'\mathbf{c}(\mathbf{a}'\mathbf{E}\mathbf{a})R_{crit}}{n}} \leq \mathbf{c}'\mu\mathbf{a} \leq \mathbf{c}'\overline{\mathbf{X}}\mathbf{a} + \sqrt{\frac{\mathbf{c}'\mathbf{c}(\mathbf{a}'\mathbf{E}\mathbf{a})R_{crit}}{n}} \qquad (6)$$

where **c**, μ, and **a** are as previously defined. $\bar{X}$ is a k ×p matrix of sample means, n is the number of subjects per group, and R_{crit} is the α-level critical constant for the R statistic of $\mathbf{HE}^{-1}$ (Harris, 1985). In this way, all possible contrasts of the type, $\mathbf{c}' \bar{X} \mathbf{a}$, can be used to construct intervals, of which 1-α of these intervals will include the population multivariate contrast, $\mathbf{c}'\mu\mathbf{a}$.

Gabriel (1968) extended the multivariate STP's to the other multivariate test statistics: W, V, and T. Gabriel (1968) also determined the critical constants for simultaneous tests made of minimal hypotheses; single linear parametric functions or univariate contrasts. Gabriel defined the critical constants for minimal hypotheses on the R, V, W, and T STP's as:

$$R = \frac{R_\alpha}{1-R_\alpha}; \quad T^2 = \frac{T_\alpha^{\,2}}{\upsilon_e}; \quad V = \frac{V_\alpha}{1-V_\alpha}; \quad W = W_\alpha^{-1} - 1. \tag{8}$$

When p = 1, each is equivalent to the Scheffe` critical constant; $\frac{\upsilon_h}{\upsilon_e} F_\alpha^{\,1}$ (Bird & Hadzi-Pavlovic, 1983). When $s > 1$ the MANOVA STP critical constants vary, and the R critical constant will be less than the others. Hence, Gabriel concluded that the R STP is the most resolvent STP; it will reject more hypotheses than the other STP's. All of these STP's are coherent with the corresponding omnibus test, but only the R STP is also consonant with the corresponding omnibus test. This follows from the observation that the R statistic tests the population of contrasts of the greatest root. Whereas the W, V, and T statistics test the population contrasts on the combined *s* roots. Therefore, when discussion of STP's is restricted to follow-up tests of the greatest root, only the R STP sample space is being tested. If contrasts on the remaining *s* - 1 roots were considered, all of the sample space of the V, W, or T statistics would be included. In this case, the V, W, and T statistics would have both the properties of coherence and

consonance. Most comparisons of the R, V, W, and T STP's have only been concerned with follow-up tests on the greatest root (Bird & Hadzi-Pavlovic; Gabriel). The greatest root often has the most practical significance and is generally of most concern to researchers. Therefore, the R STP can be expected to provide the greatest power for the most relevant follow-up questions. However, to ensure the property of coherence, the STP must be conducted with the same test statistic as for the overall test. Olson (1974, 1976) recommended the V statistic for general use due to its robustness to different assumption violations. Additionally, the W and the T test statistics are still widely used. Therefore, the R STP is not always the most appropriate STP, even though it is the most resolvent on follow-up tests of the first discriminant function.

Although, multivariate simultaneous test procedures have been criticized for lacking sufficient power, Barcikowski and Elliott (1991) have shown that this is due to the limited circumstances under which they have been used. It has been demonstrated that the power of SCI's/STP's can increase dramatically when few restrictions are placed on the dependent variable set (Bird & Hadzi-Pavlovic, 1983). Elliott (1993) also found that R SCI's had power close to the omnibus MANOVA test under certain circumstances.

Moderately-restricted contrasts

Multivariate contrasts can be completely unrestricted, in which the linear combination of the dependent variable set that maximally separates some linear combination of the groups is identified. For instance, using data from Wilkonson (1975), Barcikowski and Elliott (1991) determined that the composite variate of the three dependent variables which maximally separated the groups was equal to $V1 = .44Y_1 - .79Y_2 - Y_3.43$. Therefore the **a** vector for this composite was

a = {.44, -.79, -.43}. The linear combination of groups that the **a** vector maximally separated can also be determined. The contrast coefficients for the groups are contained in the **c** vector. In this instance, they were $-.72\mu_1 + .70\mu_2 + .02\mu_3$, therefore the **c** vector was **c** = {-.72, .70, .02} (Barcikowski & Elliott, 1991). If this **a** vector and **c** vector were used to create a Roy-Bose interval, the interval would be consonant with the omnibus test; the unrestricted Roy-Bose contrast would not contain the hypothesized population parameter if the omnibus test was significant.

Conversely, strong restrictions could be placed on the contrast coefficients such that only univariate comparisons of pairs of groups are tested. By simplifying the **a** and **c** vectors above, a strongly-restricted contrast could be formulated, such as **a** = {0, 1, 0} and **c** = {1, -1, 0}. This would be a contrast of the first and second group on the second dependent variable. This type of restriction simplifies the contrast to a very interpretable univariate analysis. However, strongly-restricted contrasts have very low power (Barcikowski & Elliott, 1991).

Moderately-restricted contrasts are a compromise between interpretability and power. The unrestricted vectors above suggest the contrast, **a** = {1, -1, 1} and **c** = {-1, 1, 0}. This would be a contrast of the difference of the combination of variables one and three with variable two between the first and second groups. This contrast has more power than the strongly-restricted contrast and is still reasonably interpretible.

Power and Robustness of Moderately-restricted Contrasts

The power and robustness of multivariate simultaneous test procedures involving moderately-restricted contrasts has only been investigated in two studies. Bird and Hadzi-Pavlovic (1983) compared the V and R STP's for a one-factor

MANOVA with 36/k subjects in each group, $\alpha = .05$, and a noncentrality structure that was diffuse across the s roots. They varied group size, the number of dependent variables, inter-variable correlations, and level of contamination of heterogeneity of variance-covariance heterogeneity. They studied unrestricted contrasts, moderately-restricted contrasts, and strongly-restricted contrasts. Elliott (1993) investigated the R SCI under various conditions, with and without assumption violations, when contrasts were moderately-restricted. Elliott's study investigated whether the conservative effect of moderately restricting contrast coefficients balanced out the liberal effect of the violation of variance-covariance heterogeneity, and was adequate to ensure robustness in most situations. Elliott investigated the power and robustness of the R SCI following a significant omnibus test in one-way MANOVA with equal n's with varying numbers of dependent variables, numbers of groups, α-levels, three types of noncentrality structures, with violations of the normality assumption and the homogeneity of variance-covariance assumption. Fixed conditions of the study included: effect size (ES) = .5; power = .8 or .9; and moderate restrictions of the type of contrasts made.

These two simulation studies that investigated the power and robustness of MANOVA STP's/SCI's found patterns similar to what Olson (1974, 1976) found for omnibus tests (Bird & Hadzi-Pavlovic, 1983; Elliott, 1993). Kurtosis usually had a conservative effect on the R SCI/STP; reducing actual α below that of nominal α, and reducing the empirical power of the R SCI/STP relative to the omnibus test. Heterogeneity of variance-covariance matrices had a liberal effect on the Type I error rates of the V STP and R STP/SCI. In some cases, the exceedance rates reached unacceptable levels. Differing effects of heterogeneity of variance-covariance matrices on power were found. Bird and Hadzi-Pavlovic demonstrated that increasing restrictions on the contrast

coefficients had a conservative effect on the STP's. However, Elliott generally found that the increased conservativeness on the Type I error rates of the R SCI due to imposing moderate restrictions on the contrasts was not enough to counterbalance the liberal effect of introducing heterogeneity of variance-covariance.

These findings fail to identify an optimal STP to use for coherent MANOVA follow-up tests, when violations of multivariate assumptions might be suspected. The V STP is too conservative to be of any practical use when one wishes to make easily interpretable contrasts. The robustness of the R STP/SCI to heterogeneity of variance-covariance has not been resolved. Although Bird and Hadzi-Pavlovic's (1983) findings appeared to indicate that imposing moderate restrictions on the types of contrasts made might negate the liberal effect of violating the assumption of heterogeneity of variance-covariance matrices, Elliott's (1993) study did not confirm this. Elliott's results also suggested that the power may be reduced to inadequate levels by violating this assumption. Based on Olson's (1974, 1976) findings comparing the robustness and power of all four of the omnibus test statistics; V, W, T, and R, it can be inferred that the power and robustness of the W and T STP's are probably intermediate between the V and R STP's.

Therefore, the purpose of this study is to compare the power and robustness of V, W, T, and R STP's using moderately-restricted contrasts with and without violations of multivariate assumptions. By doing so, this study should help to determine which multivariate test statistic would yield the best compromise of power and robustness in STP's, under different conditions.

Methodology

Monte Carlo simulation methods were used to compare the robustness and power of the STP's of the four commonly used MANOVA test statistics: W, V, T, and R. This comparison among the four STP's was made with and without violations of MANOVA assumptions. The power and robustness of all four of the STP's was compared on the first discriminant function variate.

Simulation Design

Monte Carlo Technique

Monte Carlo simulation was used to generate multiple samples from a population with a known covariance structure and centrality or noncentrality structure. A SAS-IML program was created to set the population parameters and randomly generate the sample data.

The number of replications was determined from Barcikowski and Robey's (1988) table of iterations needed for Monte Carlo studies. Liberal estimates of the number of iterations necessary to maintain the actual α-level within .25 of the nominal α-level of .05 is 5042 replications. Accordingly, 6000 replications of each combination of conditions were simulated.

Conditions Modeled

Population structure.

If **F** is the parametric analog of the **H** matrix (2), and **V** is the parametric analog of the **E** matrix (3), then $\mathbf{G} = \mathbf{F}\mathbf{V}^{-1}$ is the parameter estimated by the $\mathbf{H}\mathbf{E}^{-1}$ matrix. Hence, $\mathbf{H}\mathbf{E}^{-1}$ is a statistic estimating the parameter **G**. The **F** matrix can take on an infinite number of forms in the noncentral case. The specific noncentrality structures used in this study are described in the "noncentrality structure" section. The covariance matrix, **V**, can also take on an infinite number of forms. However, **V** can be simplified, because it is a positive definite matrix; a symmetric matrix with positive eigenvalues. For every positive definite matrix

there exists an orthogonal matrix, **C**, such that **C'VC** = **I.** Further, the test criteria are functions of the eigenvalues, and are not affected by translations, rotations, or scale changes of the axes (Anderson, 1958, p. 221-224). Hence irrespective of the correlation structure among the dependent set of variables, the covariance matrix can be reduced to the identity matrix. Therefore, **I** was used as the covariance matrix when MANOVA assumptions were met.

Noncentrality structure.

Noncentrality was introduced in four ways. The noncentrality structure was either concentrated in one characteristic root or diffused across the *s* roots. Two types of concentrated structures, C1 and C2, and one type of diffuse structures, D1, were created. The C1, C2, and D1 structures were equivalent to the noncentrality structures termed Type 1, Type 2, and diffuse, respectively, in previous research (Elliott, 1993; Olson, 1974).

The three types of noncentrality structures were constructed as follows.

(1) C1 was constructed with the population mean vector of group 1 = $\mu_1 = \{kc_1, kc_2, \ldots kc_p\}$ and with the null vector for all other groups. Hence, group 1 differed from all other groups on all p variables. The constant, c, is a constant chosen to produce a specified noncentrality parameter. The resulting eigenvalues of the population **G** matrix are: $(pnk(k-1)c^2, 0, \ldots)$, where p is the number of dependent variables, k is the number of groups, and n is group size (Olson, 1974).

C1

$$\begin{bmatrix} kc_{11} & kc_{12} & \cdots & kc_{1p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \quad (8)$$

(2) The population mean vector of group 1 in the C2 structure was $\mu_1 = \{kc_1, 0, \ldots 0_p\}$, while the null vector was used for all other groups. Therefore group 1 differed from all other groups only on variable 1. The resulting eigenvalues of the population **G** matrix are: $(nk(k-1)c^2, 0, \ldots)$ (Olson).

C2

$$\begin{bmatrix} kc_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{9}$$

(3) In diffuse structure, D1, there are group mean differences in all dimensions of the *s*-space. All elements of each group vector are set equal to zero, except the ith element of the ith group mean, which was set equal to kc for all $i \leq s$. Therefore, group 1 differed from all other groups on variable 1, and group two differed from all others on variable 2, and so on. The resulting eigenvalues of the **G** matrix depend on whether $s = p$ or $k - 1$. When

$s = p$, there are $p - 1$ roots equal to nk^2c^2 and one root equal to $nk(k - p)c^2$. When $s = k - 1$, there are $k - 1$ roots equal to nk^2c^2 and the remaining $p - s$ roots are necessarily equal to zero (Olson).

D1

$$\begin{bmatrix} kc_{11} & 0 & \cdots & 0 \\ 0 & kc_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & kc_{kp} \end{bmatrix} \tag{10}$$

Noncentrality parameter and effect size.

The noncentrality parameter, NCP, was measured as the sum of the eigenvalues of the **G** matrix when MANOVA assumptions were met. The noncentrality parameter was varied to maintain a moderate effect size; $f^2 = .15$ (Cohen, 1988, p.480). The noncentrality parameter was related to the effect size by the equivalency: $NCP = f^2(u + v + 1)$ (Cohen, p.481), where u = numerator df and v = denominator df (Cohen, p.471). The values of the noncentrality parameters used for each combination of p and k, to maintain effect size at .15, are given in Table 3. For example, when p=2, k=3, and noncentrality structure=C_2, group 1 would need to be 1.16 standard deviations greater than the other two groups on variable 1 to generate this level of effect size.

Power and n-size.

The power of the omnibus test was maintained at .8. Cohen's (1988) power calculations were based on Wilks' lambda. However, Olson's (1974) results suggested that the power levels of the W, V, T, and R test statistics are close, when MANOVA assumptions are met. This power level was fixed high enough to allow for the reduction of power that occurs when STP's of restricted contrasts were formulated. Yet, this power level still allows for some fluctuation among the test criteria. Sample size, n, was determined by the procedure given by Cohen (1988, p. 515) for calculating n-size of set correlations.[2] All groups had equal n-size. The sample sizes used for each level of k and p, to maintain power at .8, are given in Table 3. The power charts were not given for $\alpha = .10$, therefore the same values derived for $\alpha = .05$ were used for $\alpha = .10$.

Number of dependent variables.

The number of dependent variables, p, simulated in this study was two, four, and six. Belli (1989) found that 70% of the one-way MANOVA analyses recently published in American Educational Research Journal (AERJ) during a

five-year period used p = 4. Hence, the number of dependent variables simulated in this study bracketed p = 4.

Number of groups.

This study investigated group sizes, k, of three and four. Bird and Hadzi-Pavlovic (1983) recommended that the results for both k = 4 and k = 6 not be examined in detail, presumably because the patterns of difference between R and V STP's were similar for both. Belli (1989) found that the most common number of groups investigated in recent studies published in AERJ was two. The next most common group size was four, and the largest number of groups studied was five. This study did not simulate groups as small as 2, because it is not necessary to make contrasts across 2 groups. However, the group sizes simulated in this study were feasible values according to recent research.

Alpha level.

The α-level used for significance criteria for the four STP's was .05.

Violations of Distributional Assumptions Modeled

Introducing contamination.

To introduce contamination into the covariance structure in order to model violations of MANOVA assumptions, the contaminated normal distribution was used (Andrews, 1972, p.57-61). Olson (1974) generalized this procedure of adding contamination to multivariate applications. Olson demonstrated:

> if **Q** (p x p) equals $\mathbf{V_1}^{-1/2}$ with probability (1 - t), and equals $\mathbf{V_2}^{-1/2}$ with probability t, then the random vector **Y** (p x 1) = $\mathbf{Q^{-1}Z}$ has a contaminated normal distribution such as would result from sampling with probability (1 - t) from the p-variate population N(0, $\mathbf{V_1}$) and with probability t from N(0, $\mathbf{V_2}$) for any population covariance matrices $\mathbf{V_1}$ and $\mathbf{V_2}$, where **Z** (p x 1) is a vector of independent standard normal deviates. (p. 895).

Therefore, a mixture of N(0, $\mathbf{V_1}$) and N(0, $\mathbf{V_2}$) can be reduced to a mixture of

N(0, **I**) and N(0, **D**). An analogous situation exists for the noncentral case (Olson). In this study, the uncontaminated population was distributed as N(0, **I**) and the contaminated population was distributed as N(0, **D**) in the null case.

Type of violation.

Two types of violations of distributional assumptions were modeled in this study: violation of the assumption of multivariate normality in the form of kurtosis and heterogeneity of variance-covariance matrices.

Kurtosis was introduced mainly to investigate whether the power of the W, V, T, and R STP's was still adequate, under varying conditions, when kurtosis was present. Of particular interest was "thick-tailed" distributions (platykurtic), in which there were many observations with extreme scores from the mean. These distributions commonly cause inflated estimates of error variance and inaccurate parameter estimates (Judd & McClelland, 1989, p. 210). "Thin-tailed" distributions (leptokurtic) cause very little data-analytic problems (Judd & McClelland, p. 499). Therefore, only kurtosis in the form of platykurtic distributions was addressed in this study. The method of adding kurtosis was the same as was used by Olson (1974) and Elliott (1993). Using Olson's notation kurtosis was introduced in the form of $(a_1, a_2, \ldots, a_k)$, where a_1 was the proportion of observations in group 1 drawn from a distribution with higher variability. Therefore, all groups were equally affected. In this way, only kurtosis and not heterogeneity of variance-covariance matrices was introduced. In this study, each a_i was set equal to .20. Olson found that consequences of kurtosis were most serious when a_i was equal to .10 or .20 as opposed to values of a_i equal to .02 or .40.

Heterogeneity of variance-covariance matrices was introduced primarily to study its effect on the Type I error rates of the W, V, T, and R STP's. The current study was designed to determine if any of the STP's produced acceptable Type I

error rates in the presence of heterogeneity of variance-covariance matrices when contrasts were slightly or moderately restricted. Heterogeneity of variance-covariance matrices was added by the method used by Olson (1974) and Elliott (1993). As previously stated, heterogeneity of variance-covariance matrices can arise from differing intervariable correlations among the k groups, or from heterogeneity of variance for any of the dependent variables. The method used in this study introduced heterogeneity of variance-covariance matrices with violations of homogeneity of variance. Using Olson's notation, heterogeneity of variance-covariance matrices of the form (a_1, 0, 0, . . .) was introduced, where a_i was equal to one. Therefore, the heterogeneity of variance-covariance matrices was concentrated in one group, in which 100% of its observations came from a distribution of higher variability. Olson found that patterns that included both kurtosis and heterogeneity of variance-covariance matrices produced effects intermediate between these two extremes. An example of this intermediate pattern would be when 40% of the observations in group 1 only came from a distribution with larger variance, Consequently, only the extreme situations were modeled in this study.

Concentration of contamination.

This factor refers to how the contamination was introduced relative to the dependent variable set. Following the method of Olson (1974), two levels of concentration of contamination were used. In the low-level of concentration of contamination, all dimensions of the p-space were equally contaminated, such that the contaminating covariance matrix was $\mathbf{D} = d\mathbf{I}$, where d was the degree of contamination. In the high concentration of contamination condition, contamination only occurred in one dimension of p-space. The contaminating covariance matrix was $\mathbf{D} = \text{diag}(pd - p + 1, 1, 1, \ldots)$. This covariance matrix was

chosen in order to maintain the same total variability in the low-concentrated and high-concentrated conditions.

Degree of contamination.

The degree of contamination, d, indicates how much more variable the contaminating distribution was relative to the uncontaminated distribution. Olson (1974) used levels of d = 4, 9, and 36, and was subsequently criticized for using levels of contamination unrealistically high (Stevens, 1979). In this study, the degree of contamination modeled was d = 4 and d = 9.

Procedures

The general procedure followed in this study was as follows. First situations were simulated using all combinations of the conditions and assumption violations previously mentioned. The procedures to be described are given in Table 1. For each situation, omnibus tests were conducted for each of the four MANOVA test statistics, W, V, T, and R. If the omnibus test was not statistically significant, no further investigation was made of that situation with that particular test statistic. When a significant omnibus test was detected, the maximized STP contrast was generated. From this maximized contrast, further restricted contrasts were made. The type I error and power of the STP'S was determined by the method described in "Power and Robustness of the STP's".

Insert Table 1 here

Restrictions Imposed on Contrast Coefficients

The contrast coefficients used in the moderately-restricted condition were derived from the unrestricted, maximized contrasts. The unrestricted contrast was generated by calculating the eigenvector associated with the particular root of interest. Normalizing this eigenvector produced **a**, the vector of contrast

coefficients for the dependent set. **Ma** is equal to **c**, the vector of contrast coefficients for the groups. **M** is a matrix of deviation means, standardized with respect to within-group variance (Bird & Hadzi-Pavlovic, 1983).

The contrast coefficients of the dependent set used in the moderately-restricted condition were limited to values of -1, 0, and 1. To generate these coefficients the method of Bird and Hadzi-Pavlovic (1983) and Elliott (1993) was used. Each element of the **a** vector was divided by the largest value of the **a** vector, and then the fractions were rounded off to the nearest $\pm$ 1 or 0. The group contrast coefficients in the moderately-restricted condition were all (n_1, n_2) contrasts of the groups. This amounted to six contrasts for the three-group condition and 25 contrasts for the four-group situation.

Power and Robustness of the STP's

When the population had a central structure, any significant (n_1, n_2) contrasts were counted toward type I error for that STP. For instance, if two of the six possible contrasts for Roy's STP were significant in a particular replication, then the type I error for Roy's STP in that example would be .167. Type I error was then averaged over all replications for a particular simulation. Elliott (1993) determined that his simulations had poor matches of the **c** vector of the significant STP to the population structure when the population structure was noncentral. Therefore, in this study power was determined by analyzing the proportion of times the particular (n_1, n_2) contrast that fully represented the induced noncentrality structure was found to be significant. For instance, if a simulation of the C2 noncentrality structure (group one differs from all other groups on the first dependent variable) produced a significant contrast between group one and groups two and three, this would be counted toward the power of that contrast. However, if a simulation of the C2 noncentrality structure produced a significant contrast

between groups two and three, this would not be counted toward the power of the contrast.

Quality Control

To determine whether the SAS-IML program was correctly calculating the test statistics and discriminant function weights, the simulation runs were periodically selected and run on the SAS (version 6.07) PROC CANDISC.

Additionally, the calculation of Type I and Type II errors of individual contrasts allowed for comparison of the specific significant contrasts with the population structure. If the contrasts declared significant were not those imposed in the population structure, then the usefulness of the STP procedure to follow-up significant omnibus MANOVA's was questioned.

Results

Type I Error Rates

The type I error rates of moderately-restricted contrasts of the first root were conservative under all conditions investigated. The type I error rates of all the test criteria were the most inflated in the presence of a low concentration of heterogeneity of variance-covariance (see figure 2.). In this case, the STP of Roy's Greatest Root had higher type I error rates than the STP's of the other test criteria. This distinction became greater as the number of variables increased. The most conservative test statistic was the Pillai-Bartlett STP.

Insert Figures 1-3 here

Power

The pattern of power values of moderately-restricted contrasts of the three noncentrality structures often differed. However, there were some robust trends in power which were exhibited in all noncentral population structures. First, under conditions in which the test criteria diverged, Roy's STP had the greatest power, followed by the Hotelling-Lawley STP, then Wilks' STP, and lastly the Pillai-Bartlett STP. Second, power increased in the presence of heterogeneity of variance-covariance and decreased in the presence of kurtosis. Third, the power was highest when the number of variables was equal to two.

The two concentrated noncentrality structures generally had higher power values than the diffuse structure (see figures 4-12). Power levels were acceptably high when the number of variables was equal to two without assumption violations or in the presence of heterogeneity of variance-covariance (see figures 4, 5, 7, 8, 10, & 11). The test criteria diverged most when the number of variables was equal to two, assumption violations were met or kurtosis was present, and the noncentrality structure was concentrated (see figures 4, 6, 7, & 9). In these instances, Roy's STP had the largest difference in power from the other STP's. The C2 noncentrality structure had different power patterns from the other two noncentral structures under most conditions (see figures 4-12). The test criteria in this noncentrality structure did not have such a dramatic drop in power as the number of variables increased.

Insert Figures 4-12 here

Conclusions

The test properties of the moderately-restricted STP was investigated in this study, because it has been suggested that the moderately-restricted STP is a good compromise between interpretability and power (Barcikowski & Elliott, 1991; Elliott, 1993). Although, the results of this study did not support that hypothesis, some general conclusions can be made about the choice of MANOVA test statistics based on test properties of the STP. If one adopts Olson's view of type I and type II errors, high type I error rates make the test more dangerous and high type II error rates make it less useful. If the choice of the test statistic was based on test properties of the STP's, then Roy's Greatest Root would be recommended in the presence of heterogeneity of variance-covariance. All the STP's had conservative type I error rates, even in the presence of heterogeneity of variance-covariance, but Roy's STP had the least conservative type I error rates and the greatest power of all the STP's in a concentrated noncentrality structure.

Kurtosis has a conservative effect on both type I error and power. If one suspected kurtosis or wanted to protect against it, the choice of the test statistic would probably be based on power, since all the test criteria have conservative type I error rates in follow-up STP's. The results of this study suggest Roy's Greatest Root would also be the recommended test statistic any in the presence of kurtosis, even in a diffuse noncentrality structure.

The moderately-restricted contrast proved to be too conservative to be very useful unless the noncentrality was concentrated in one group and one variable or the number of variables was limited to two. Therefore, the moderately-restricted contrast is not an optimum middle-ground in the sequence from the totally unrestricted contrast, which is consonant with the omnibus test for Roy's Greatest Root, to contrasts among the groups on one variable, which is very conservative relative to the omnibus test. The results of this study indicate that although the

moderately-restricted contrast is robust in terms of type I error, it lacks sufficient power in most situations.

References

Anderson, T. W. (1958). An introduction to multivariate statistical analysis. New York: Wiley.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). Robust estimates of location: Survey and advances. Princeton: Princeton University Press.

Barcikowski, R. S., & Elliott, R. (1991, October). Exploratory multivariate analysis of variance: Contrasts and variables. Paper presented at the annual meeting of the MidWestern Educational Research Association, Chicago, IL.

Barcikowski, R. S., & Robey R. (1988, April). The number of iterations in Monte Carlo studies of robustness. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. Proceedings of the Cambridge Philosophical Society, 35, 180-185.

Belli, G. M. (1989, March). Parameters in MANOVA robustness studies: Comparability to empirical research. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Bird, K. D. (1975). Simultaneous contrast testing procedures for multivariate experiments. Multivariate Behavioral Research, 10, 343-352.

Bird, K. D., & Hadzi-Pavlovic, D. (1983). Simultaneous test procedures and the choice of a test statistic in MANOVA. Psychological Bulletin, 93(1), 167-178.

Bock, R. D. (1966). Contributions of multivariate statistical methods to educational research. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology (pp. 820-840). Skokie, IL: Rand McNally.

Bray, J. H., & Maxwell, S. E. (1985). Multivariate analysis of variance. Beverly Hills: Sage.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates.

Elliott, R. (1993). Contrast selection in post hoc multivariate analysis. Doctoral dissertation, Ohio University-Chillicothe.

Emmons, N. J., Stallings, W. M., & Layne, B. H. (1990, April). Statistical methods used in American Educational Research Journal, Journal of Educational Psychology, and Sociology of Education from 1972 through 1987. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Gabriel, K. R. (1968). Simultaneous test procedures in multivariate analysis of variance. Biometrika, 55(3), 489-504.

Harris, R. J. (1985). A primer of multivariate statistics (2nd ed.). Orlando, FL: Academic Press, Inc.

Hotelling, H. (1931). The generalization of student's ratio. Annals of Mathematical Statistics, 2, 360-378.

Ito, K. (1969). On the effect of heteroscedasticity and non-normality upon some multivariate test procedures. In P. R. Krishnaiah (Ed.), Multivariate analysis: Vol. 2 (pp. 87-120). New York: Academic Press, Inc.

Ito, K., & Schull, W. J. (1964). On the robustness of the T_o^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. Biometrika, 51, 71-82.

Judd, C. M., & McClelland, G. H. (1989). Data analysis. Orlando, FL: Harcourt Brace Jovanovich, Inc.

Korin, B. P. (1969). On testing the equality of k covariance matrices. Biometrika, 56, 216-218.

Lawley, D. N. (1939). A generalization of Fisher's z test. Biometrika, 30, 180-187. Corrections in Biometrika, 30, 467-469.

Maxwell, S. E. (1992). Recent developments in MANOVA applications. In Advances in Social Science Methodology: Vol. 2 (pp. 137-168). Greenwich, Conn.: JAI Press.

McKay, R. J., & Campbell, N. A.. (1982). Variable selection techniques in discriminant analysis I. Description. British Journal of Mathematical and Statistical Psychology, 35, 1-29.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association, 69, 894-908.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 83(4), 579-586.

Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. Annals of Mathematical Statistics, 26, 117-121.

Roy, J. (1958). Step-down procedure in multivariate analysis. Annals of Mathematical Statistics, 29, 1177-1187.

Roy, S. N. (1945). The individual sampling distribution of the maximum, the minimum, and any intermediate of the p-statistics on the null hypothesis. Sankhya, 7, 133-158.

Roy, S. N., & Bose, R. C. (1953). Simultaneous confidence interval estimation. Annals of Mathematical Statistics, 24, 513-536.

Schatzoff, M. (1966). Sensitivity comparisons among tests of the general linear hypothesis. Journal of the American Statistical Association, 61, 415-435.

Stevens, J. P. (1979). Comment on Olson: Choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 86, 355-360.

Stevens, J. P. (1986). Multivariate statistics for the social sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Tatsuoka, M. M. (1989). Multivariate analysis (2nd ed.). New York: Macmillan, Inc.

Wilkinson, L. (1975). Response variable hypotheses in the multivariate analysis of variance. Psychological Bulletin, 82(3), 408-412.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. Biometrika, 24, 471-494.

Table 1

Conditions Simulated

Condition	Levels Investigated
MANOVA test criteria	W, V, T, and R
noncentrality structure	central distribution and C1, C2, and D1 noncentral structures
effect size	$f^2 = .15$
power	.80 (for omnibus test)
number of dependent variables	2, 4, and 6
number of groups	3 and 4
alpha level	.05
type of contamination	kurtosis and heterogeneity of variance-covariance matrices
concentration of contamination	low and high
degree of contamination	d=1, d=4, and d=9

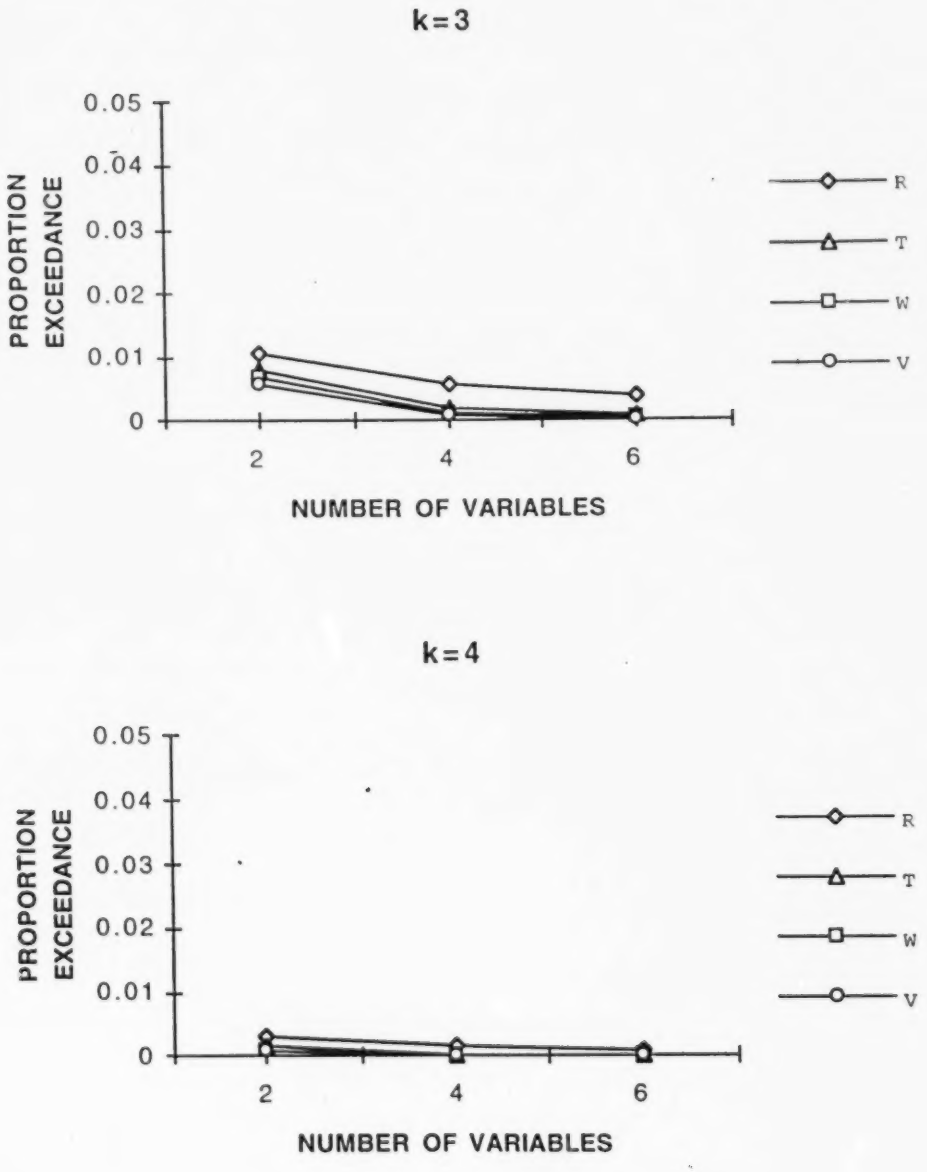

Figure 1. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root with a central structure without assumption violations; nominal $\alpha = .05$.

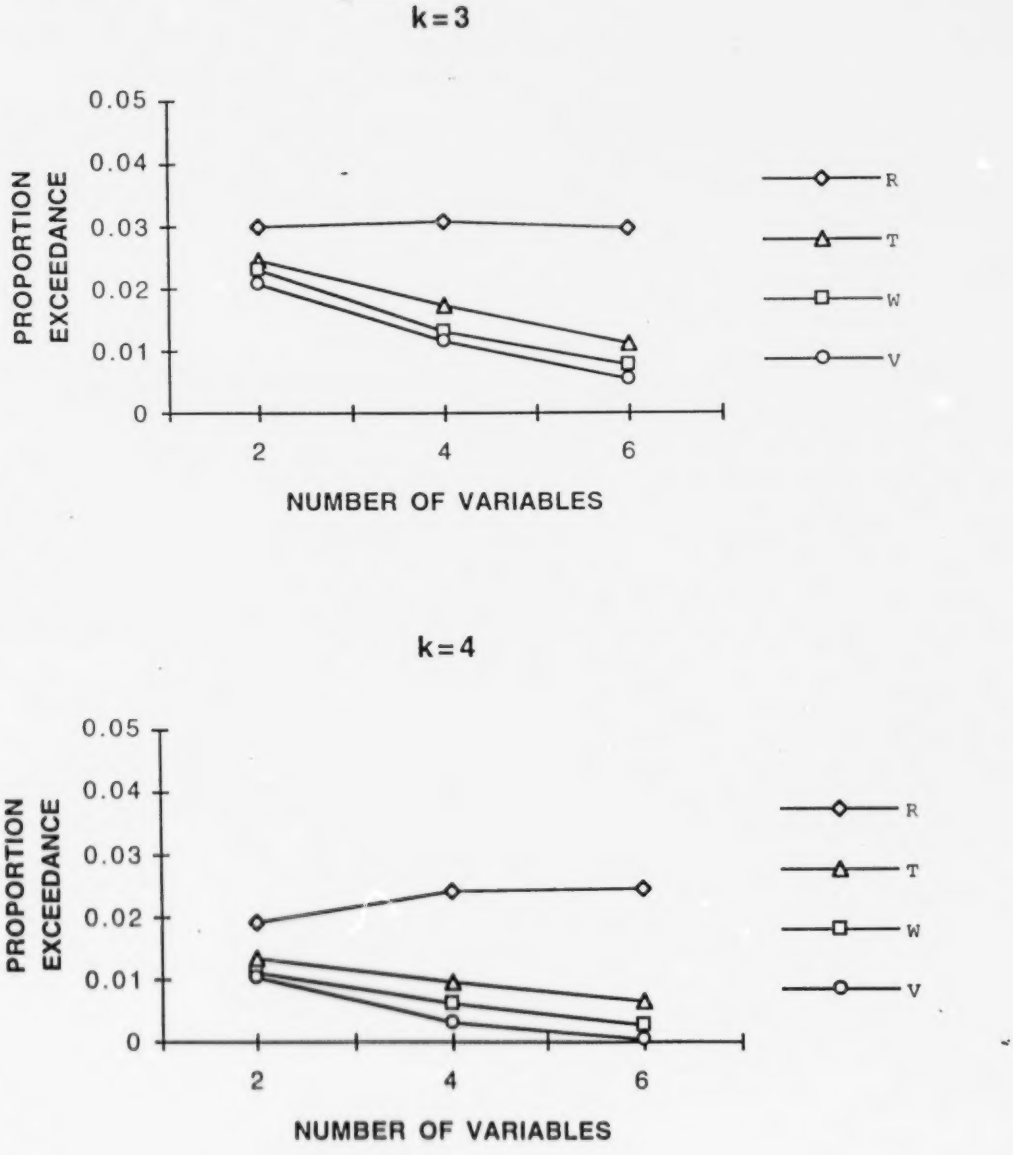

Figure 2. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root with a central structure in the presence of heterogeneity of variance-covariance when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

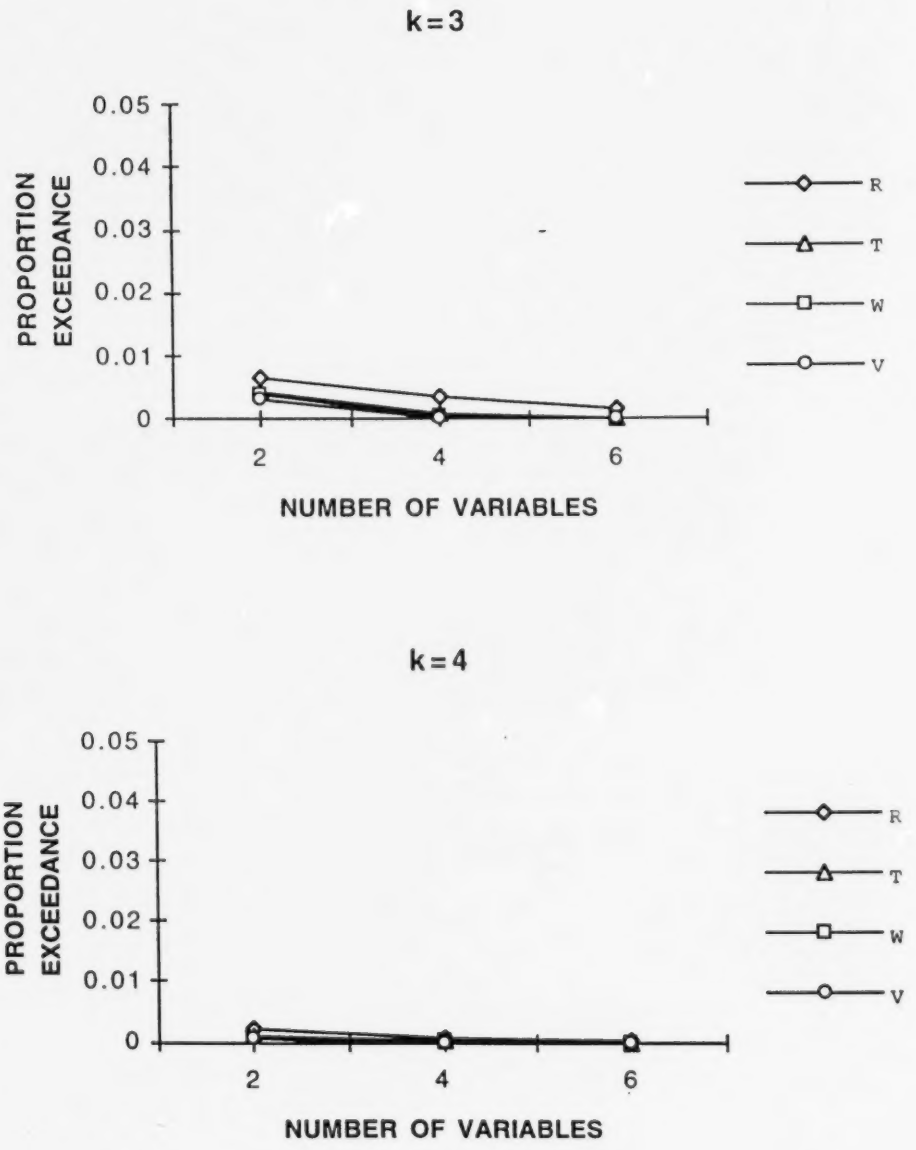

Figure 3. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root with a central structure in the presence of kurtosis when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

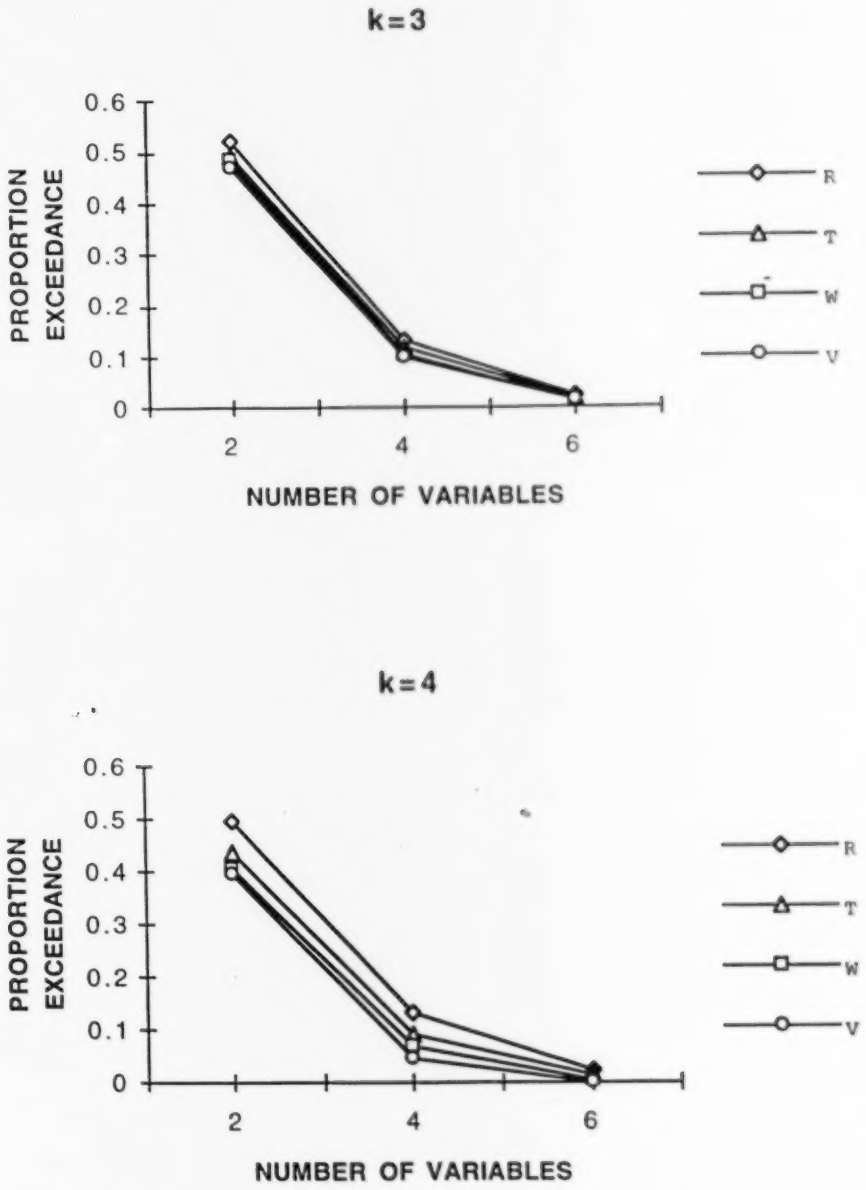

Figure 4. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C1, without assumption violations; nominal $\alpha = .05$.

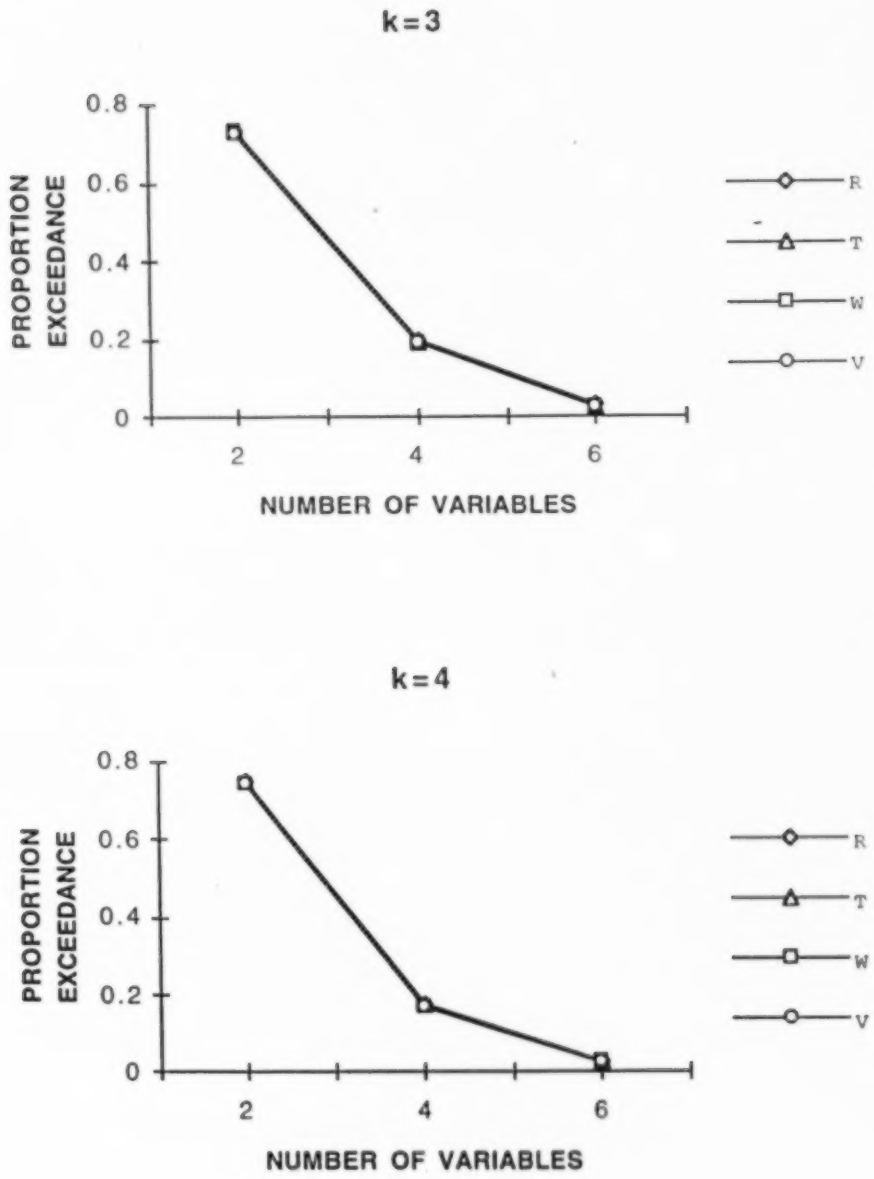

Figure 5. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C1, in the presence of heterogeneity of variance-covariance when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

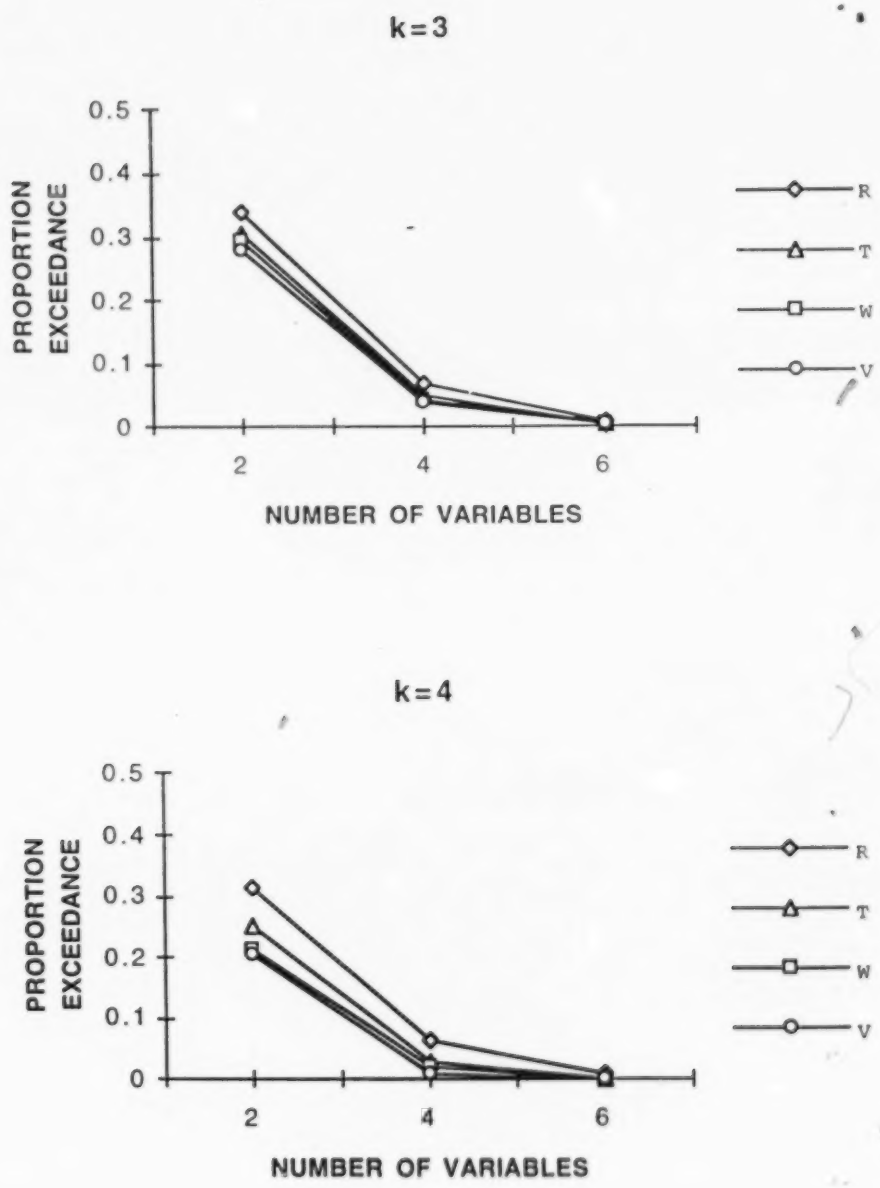

Figure 6. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C1, in the presence of kurtosis when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

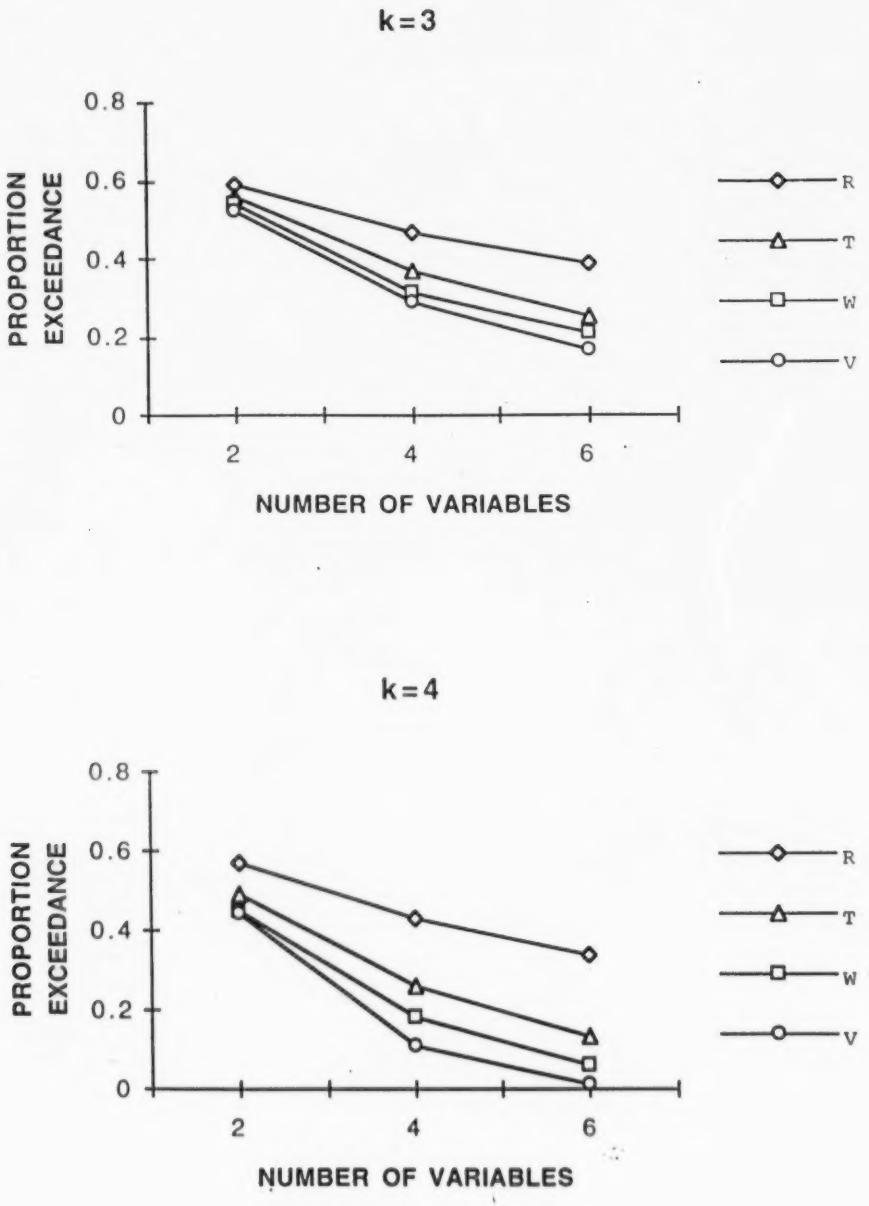

Figure 7. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C2, without assumption violations; nominal $\alpha = .05$.

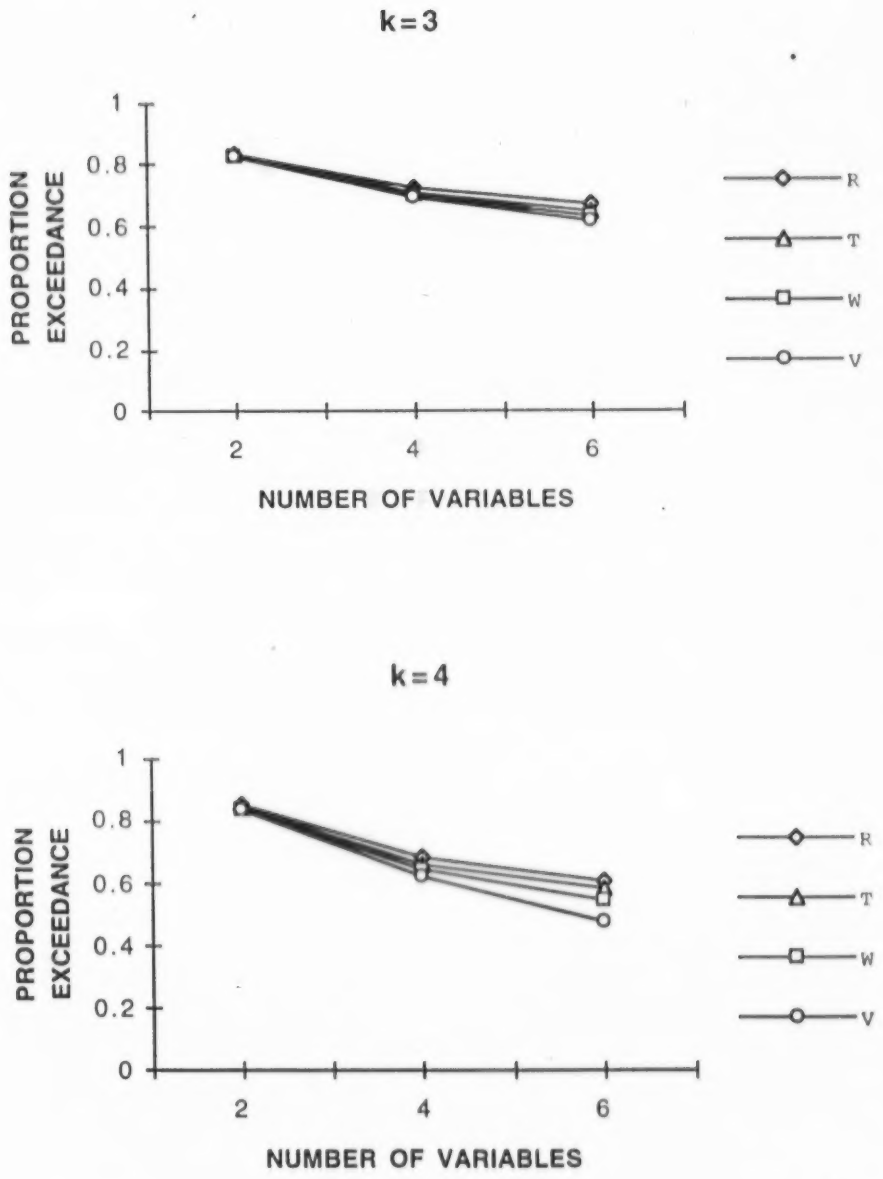

Figure 8. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C2, in the presence of heterogeneity of variance-covariance when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

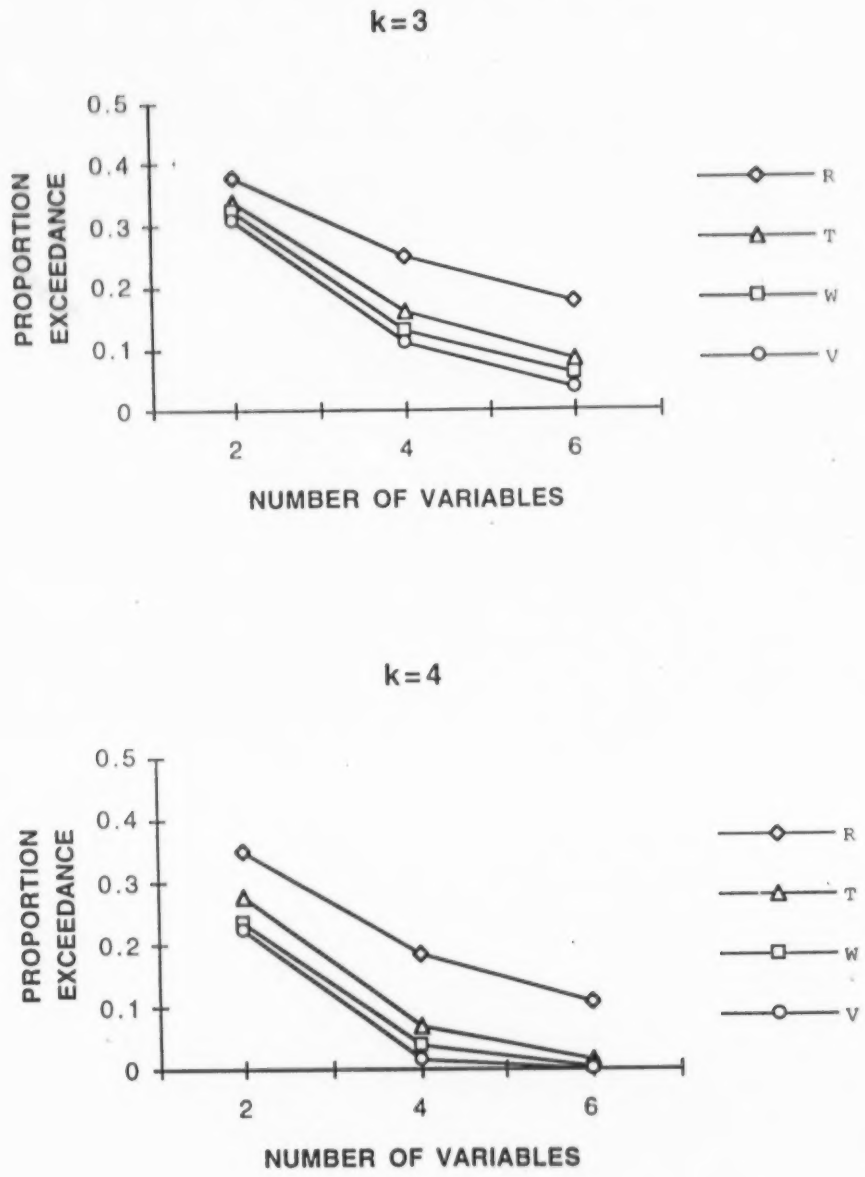

Figure 9. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for concentrated noncentrality structure, C2, in the presence of kurtosis when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.

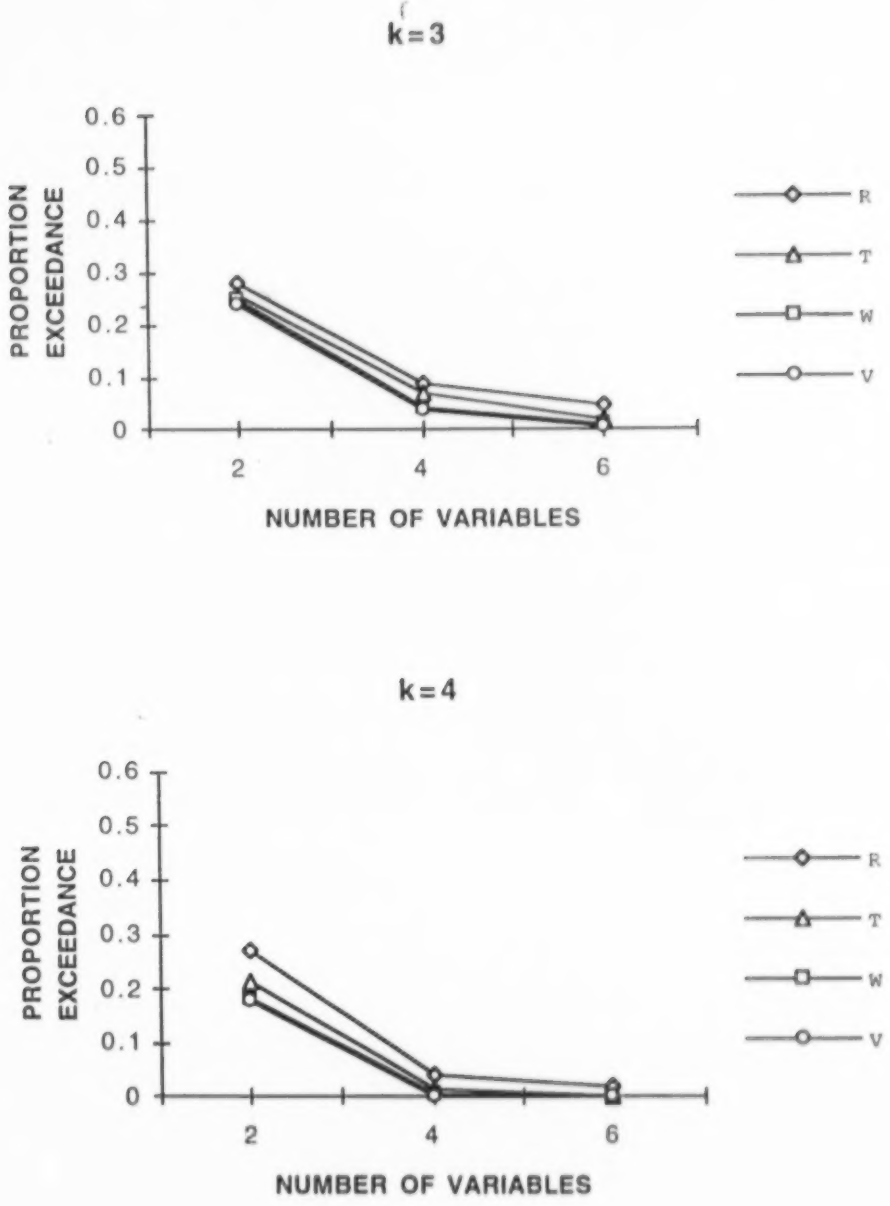

Figure 10. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for diffuse noncentrality structure, D1, without assumption violations; nominal $\alpha = .05$.

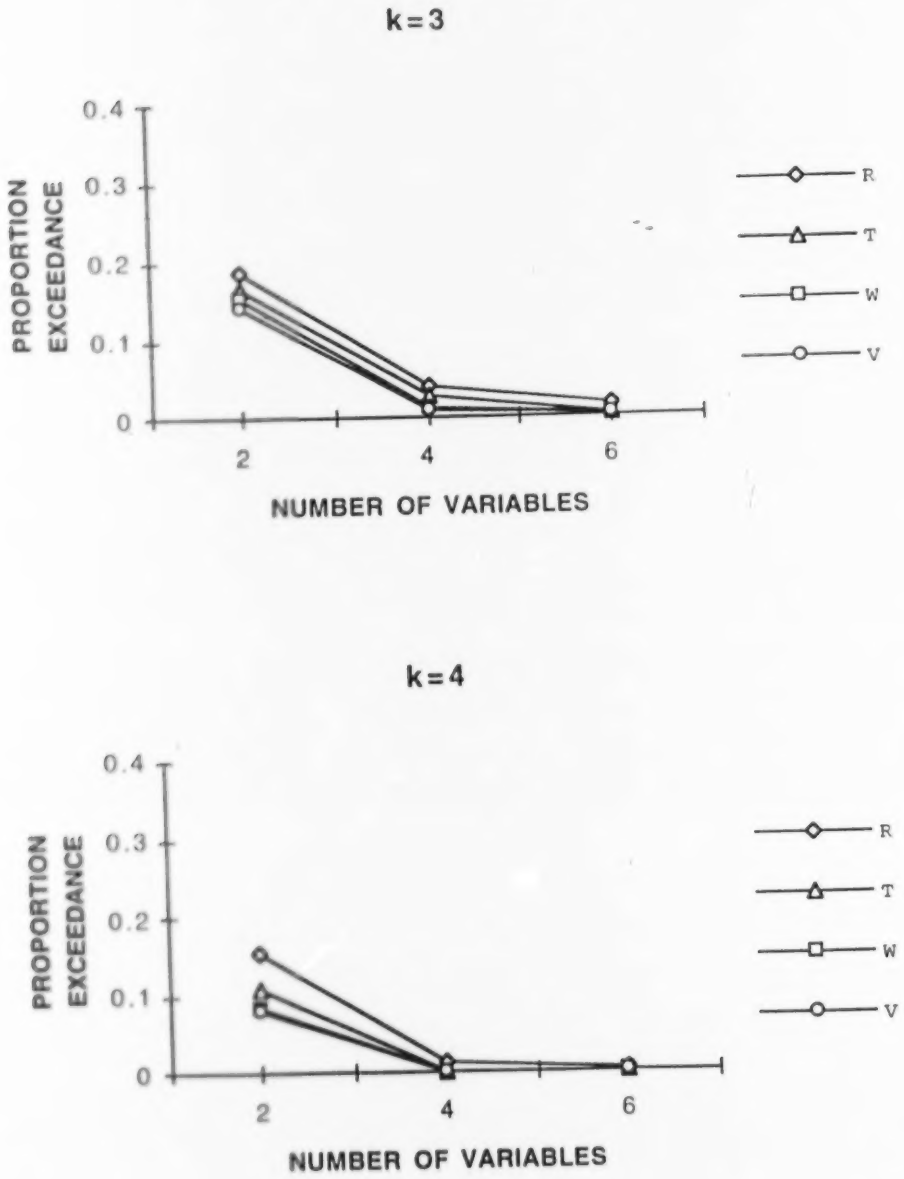

Figure 12. Proportion exceedance as a function of the number of variables for the moderately-restricted contrast of the first root for diffuse noncentrality structure, D1, in the presence of kurtosis when d=9 and the concentration of contamination is low; nominal $\alpha = .05$.